Lecture 6: MLE & Trinity of the Tests

Prepared for ECON 5033

Eric S. Lin

Department of Economics, National Tsing Hua University

November 4, 2024

Outline

- Maximum Likelihood Estimation
- Trinity of the Tests
  - Wald Test
  - Likelihood Ratio Test
  - Lagrange Multiplier Test

Motivation

- Estimation method(s) we have learned so far?
- Alternatives to OLS?
- Maximum likelihood, FIML, LIML, QMLE
- GLS, FGLS, 2SLS, 3SLS
- Instrumental variable (IV)
- Method of moment, Generalized method of moment (GMM)
- Generalized empirical likelihood, EL, CUE, ET
- Bayesian, simulation-based method, SMM, SML
- Non-parametric method, semi-parametric method

◇ will be introduced in this course

◇ will be introduced in PhD courses

Maximum Likelihood Principle

- Idea of MLE: maximize the likelihood function to obtain the estimator of the unknown parameters: R. A. Fisher's (1922) likelihood principle
- Distributional assumption will be needed
- Assume $x_1, ..., x_n$, is an i.i.d. random sample drawn from $f(x_i, \theta_o)$.
- The likelihood function (joint density) is:

$$f(x_1, x_2, ...x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \mathcal{L}(\theta | x_1, x_2, ...x_n).$$

- The MLE of θ_o is defined as

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \ln \mathcal{L}(\theta | x_1, x_2, ...x_n).$$

Maximum Likelihood Principle

Example

Suppose

$$\Pr[X_1 = x_1, X_2 = x_2, ..., X_n = x_n, \theta] = \prod_{i=1}^{n} \Pr[X_i = x_i, \theta]$$

If we happen to know the following:

$$\Pr[X_1 = x_1, X_2 = x_2, ..., X_n = x_n|\hat{\theta}_1] = 0.891$$
$$\Pr[X_1 = x_1, X_2 = x_2, ..., X_n = x_n|\hat{\theta}_2] = 0.805$$
$$\Pr[X_1 = x_1, X_2 = x_2, ..., X_n = x_n|\hat{\theta}_3] = 0.899$$
$$\Pr[X_1 = x_1, X_2 = x_2, ..., X_n = x_n|\hat{\theta}_4] = 0.811$$

The MLE of θ_o is given by:

$$\hat{\theta}_{MLE} = \hat{\theta}_3.$$

Sir Ronald Aylmer Fisher (1890 – 1962)

Maximum Likelihood Estimation

Example

Exponential distribution is given by:

$$f(x_i) = \theta e^{-\theta x_i}.$$

The likelihood function of the exponential distribution is:

$$\mathcal{L}\left(\theta | x_1, x_2, ...x_n\right) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}.$$

MLE for Linear Regression Model

- Now consider the classical linear regression model,

$$Y_i = X_i'\beta + \varepsilon_i$$

- The normality assumption of error term has been imposed
- The log-likelihood function is: (conditional likelihood)

$$\ln \mathcal{L}(\beta, \sigma^2) = -\frac{n}{2} \ln (2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta) .$$

- FOCs are

$$\frac{\partial \ln \mathcal{L}(\theta)}{\partial \beta} = \frac{1}{\sigma^2} X' (Y - X\beta) = 0$$
$$\frac{\partial \ln \mathcal{L}(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (Y - X\beta)' (Y - X\beta) = 0.$$

MLE for Linear Regression Model

- It is trivial to obtain:

$$\hat{\beta}_{MLE} = \left(X'X\right)^{-1} X'Y = \hat{\beta}_{OLS}$$

$$\hat{\sigma}^2_{MLE} = \frac{(Y - X\hat{\beta}_{MLE})'(Y - X\hat{\beta}_{MLE})}{n} = \frac{RSS}{n} = \frac{n-k}{n}\hat{\sigma}^2.$$

- The value of the log-likelihood function evaluated at the optimum is

$$\begin{aligned} &\ln \mathcal{L}(\hat{\beta}_{MLE}, \hat{\sigma}^2_{MLE}) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\frac{RSS}{n}) - \frac{n}{2}. \end{aligned}$$

- Here we notice that $\hat{\beta}_{MLE}$ is an unbiased estimator of β but $\hat{\sigma}^2_{MLE}$ is not unbiased. It shows that MLE are not always unbiased.

Properties of MLE

- Under fairly general regularity condition (See Greene, pp. 474.), MLE possesses the following nice asymptotic properties:
  1. Consistency
  2. Asymptotic normality
  3. Asymptotic efficiency
  4. Invariance
- How about the finite-sample properties?

Properties of MLE

- The MLE estimator is consistent, i.e.,

$$\text{plim } \hat{\theta}_{MLE} = \theta \text{ or } \hat{\theta}_{MLE} \xrightarrow{p} \theta$$

- In the case of linear regression model, we have

$$\text{plim } \hat{\beta}_{MLE} = \beta \text{ and plim } \hat{\sigma}^2_{MLE} = \sigma^2$$

- The MLE estimator is distributed as normality asymptotically, i.e.,

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} \mathcal{N}[0, \mathcal{I}(\theta)^{-1}],$$

where $\mathcal{I}(\theta)$ is the information matrix.

Properties of MLE

- The information matrix $\mathcal{I}(\theta)$ is defined as

$$\mathcal{I}(\theta) = -\mathrm{E}\left[\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial\theta\partial\theta'}\right] = \mathrm{E}\left[\frac{\partial \ln \mathcal{L}(\theta)}{\partial\theta}\frac{\partial \ln \mathcal{L}(\theta)}{\partial\theta'}\right]. \tag{1}$$

- Equality (1) is also known as information matrix equality under correct model specification.
- MLE estimator is asymptotically efficient in the sense that it achieves the Cramèr-Rao Lower Bound (CRLB). $\mathcal{I}(\theta)^{-1}$ provides the lower bound on the asymptotic variance-covariance matrix for any consistent asymptotically normal estimator for θ.

Properties of MLE

- Recall the information matrix inequality, we know that among all unbiased estimators, $U(X_n)$, of θ_o,

$$\mathrm{var}[\sqrt{n}U(X_n)] \geq \mathcal{I}^{-1}(\theta_o) \tag{LB1}$$

- Consider an estimator $T(X_n)$ of θ_o, which is asymptotically normal and asymptotically unbiased, i.e.,

$$\sqrt{n}(T(X_n) - \theta_o) \xrightarrow{d} \mathcal{N}(0,\ \text{asy. var}[T(X_n)]),$$

- It turns out that under some additional regularity conditions on $T(X_n)$, we can show that

$$\text{asy. var}[T(X_n)] \geq \mathcal{I}^{-1}(\theta_o). \tag{LB2}$$

Regular Estimator

Definition

A regular estimator $T(X_n)$ of θ_o which is asymptotically normal and asymptotically unbiased with asy. $\text{var}[T(X_n)] = \mathcal{I}^{-1}(\theta_o)$ is said to be asymptotically efficient.

Comments on LB1 and LB2

- Consider the MLE estimator ($\hat{\theta}(X_n) \equiv \hat{\theta}_{MLE}$) of θ_o. We know that

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta_o) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}(\theta_o))$$

- (LB1) is attained only under exceptional circumstances (i.e., usually need completeness), while (LB2) is obtained under quite general regularity conditions.
- The UMVUE tends to be unique, while asymptotically efficient estimators are not. If $T(X_n)$ is asy. efficient, then so is $T(X_n) + R_n$, provided $\sqrt{n}R_n \xrightarrow{p} 0$.
- In (LB1), the estimator must be unbiased, whereas in (LB2), the estimator must be consistent & asymptotically unbiased.
- asy. var$[T(X_n)]$ in (LB2) is an asymptotic variance, whereas (LB1) refers to the actual variance of $U(X_n)$.

Invariance Properties of MLE

- Invariance: The MLE estimator of $\gamma = g(\theta)$ is $g(\hat{\theta}_{MLE})$ if $g(\theta)$ is a continuous and continuously differentiable function.
- Note that the function $g(\cdot)$ is NOT necessarily one-to-one for the invariance property of the MLE to hold.
- For instance, MLE of β^2 is simply $(\hat{\beta}_{MLE})^2$.
- Another example is that we know MLE of p is $\hat{p} = \bar{X}$, then the MLE of

  $$\ln \frac{p}{1-p}$$

  is given by

  $$\ln \frac{\bar{X}}{1-\bar{X}}$$

- Note that proof for functions that are not one-to-one is not that obvious

Asymptotic Properties of MLE

- Asymptotic distribution of MLE?
- Need to compute $\mathcal{I}(\theta)$.
- SOCs of the original log-likelihood function are

$$\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial\beta\partial\beta'} = -\frac{X'X}{\sigma^2}$$

$$\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(Y - X\beta)'(Y - X\beta)$$

$$\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}\left(X'Y - X'X\beta\right).$$

Asymptotic Properties of MLE

- Information matrix ($\mathcal{I}(\theta) = \mathcal{I}\left(\beta, \sigma^2\right)$) is

$$\mathcal{I}(\beta, \sigma^2) = \begin{bmatrix} \frac{E(X'X)}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}.$$

- Joint asymptotic distribution is:

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{MLE} - \beta \\ \hat{\sigma}^2_{MLE} - \sigma^2 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, n\begin{pmatrix} \sigma^2 E(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}\right)$$

Properties of MLE: LB1 and LB2 Again

- Individual asymptotic distribution is trivial
- Note that var$[\hat{\beta}_{MLE}]$ attains the CRLB (in terms of LB1) but not the case for var$[\hat{\sigma}^2_{MLE}]$
- Recall that an unbiased estimator $\hat{\sigma}^2 = e'e/(n-k)$ and var$[\hat{\sigma}^2] = 2\sigma^4/(n-k)$, which is greater than $\mathcal{I}_{22}^{-1} = 2\sigma^4/n$
- In fact, no unbiased estimator of σ^2 can attain the CRLB (in terms of LB1)
- One can show:

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) = (\frac{n}{n-k})\sqrt{n}(\hat{\sigma}^2_{MLE} - \sigma^2) + \frac{k}{n-k}\sqrt{n}\sigma^2 \xrightarrow{d} \mathcal{N}(0, 2\sigma^4)$$

  - both var$[\hat{\beta}_{MLE}]$ and var$[\hat{\sigma}^2]$ attain CRLB in terms of LB2

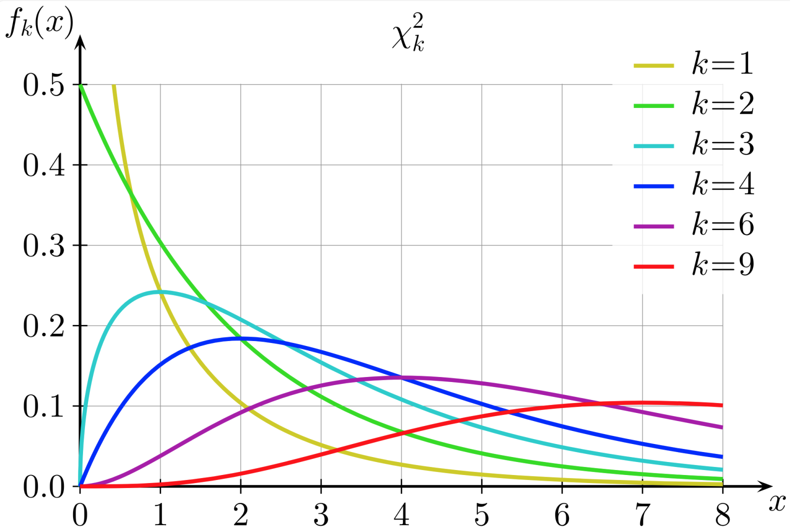
Three Classical Tests

- Consider the general test of the restriction with the form

$$R_{J\times k}\beta_{k\times 1} = q.$$

- Based on MLE, one could construct three types of classical tests.
  1. Wald test
  2. Likelihood ratio test
  3. Lagrange multiplier test
- Asymptotically equivalent tests
- The finite-sample properties are not the same though

Chi-Squre Distribution



Three Classical Tests

- We consider two sets of estimators.

$$\text{Unrestricted: } (\hat{\beta}_{MLE}, \hat{\sigma}^2_{MLE})$$
$$\text{Restricted: } (b_R, \hat{\sigma}^2_R).$$

- Note that $\hat{\beta}_{MLE} = \hat{\beta}$ (OLS) and $b_R = b^*$(RLS) but $\hat{\sigma}^2_{MLE} \neq \hat{\sigma}^2$.
- $\hat{\sigma}^2_R = (Y - Xb_R)'(Y - Xb_R)/n$ is the MLE of σ^2 under restricted model, i.e., imposing the restriction $R\beta = q$.

Wald Test

- Wald test is the most common test in econometrics and statistics.
  - t test is a special case of Wald test
  - weighted quadratic form
- It is easy because we use only unrestricted estimator.
- The basic idea is to check if the difference $R\hat{\beta}_{MLE} - q$ is close to zero.
- Recall in Lec #4 we have,

$$(R\hat{\beta} - q)'(\sigma^2 R \left(X'X\right)^{-1} R')^{-1}(R\hat{\beta} - q) \sim \mathcal{X}^2 (J) . \tag{2}$$

- Idea is to replace the unknown parameter σ by sample counterpart.

Wald Test

- We also have

$$F = \frac{(R\hat{\beta} - q)'(\hat{\sigma}^2 R\,(X'X)^{-1}\,R')^{-1}(R\hat{\beta} - q)}{J} \sim F(J, n-k). \quad (3)$$

- One can construct the usual Wald type test statistic,

$$W = \frac{(R\hat{\beta} - q)'(R\,(X'X)^{-1}\,R')^{-1}(R\hat{\beta} - q)}{\hat{\sigma}^2} = JF \overset{d}{\sim} \mathcal{X}^2\,(J)\,.$$

- More precisely, we should replace $\hat{\sigma}^2$ by $\hat{\sigma}^2_{MLE}$ to obtain $W = \frac{n}{n-k} JF$.
- The usual t-test is nothing but a Wald test ($J = 1 \Rightarrow Wald = t^2$)

Likelihood Ratio Test

- Likelihood ratio test is also known as LR test.
- Both restricted and unrestricted models should be estimated.
- Idea is to check whether the difference in log-likelihood values $\ln \mathcal{L}(\hat{\theta}_R) - \ln \mathcal{L}(\hat{\theta}_U)$ is significantly different from zero.
- Null hypothesis again:

$$H_o : R\beta = q$$

- The LR test statistic is:

$$-2\left(\ln \mathcal{L}_R - \ln \mathcal{L}_U\right) \overset{d}{\sim} \mathcal{X}^2\left(J\right). \tag{4}$$

- One could manipulate (4) to see the link between W and LR tests.

Likelihood Ratio Test

$$
\begin{aligned}
& -2(\ln \mathcal{L}(b_R, \hat{\sigma}_R^2) - \ln \mathcal{L}(\hat{\beta}_{MLE}, \hat{\sigma}^2_{MLE})) \\
&= -2\left(-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{RSS_R}{n}\right) + \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln\left(\frac{RSS_U}{n}\right)\right) \\
&= n\ln\left(\frac{RSS_R}{RSS_U}\right) \\
&= n\ln\left(1 + \frac{RSS_R - RSS_U}{RSS_U}\right) \\
&= n\ln\left(1 + \frac{J}{n-k}\frac{(RSS_R - RSS_U)/J}{RSS_U/(n-k)}\right) \\
&= n\ln\left(1 + \frac{JF}{n-k}\right) \simeq JF \quad [\lim_{a\to 0}\frac{\ln(1+a)}{a} = 1]
\end{aligned}
$$

Lagrange Multiplier Test

- We call it LM test or Score test.
- It requires only computing the restricted set of estimates.
- Idea is to check whether the Lagrangean multiplier is significantly different from zero.
- One would solve for the restricted MLE using Lagrange method.

$$\begin{aligned}\ln \mathcal{L}_R(\beta, \sigma^2) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) \\ & + \lambda'(R\beta - q)\end{aligned}$$

- Recall that the optimal λ under restricted LS is

$$\lambda^* = (R\left(X'X\right)^{-1}R')^{-1}(q - R\hat{\beta}).$$

Lagrange Multiplier Test

- FOCs are:

$$\frac{\partial \ln \mathcal{L}_R(\beta,\sigma^2)}{\partial \beta} = -\frac{1}{2\hat{\sigma}_R^2}(-2X'Y + 2X'Xb_R) + R'\hat{\lambda} = 0$$

$$\frac{\partial \ln \mathcal{L}_R(\beta,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}_R^2} + \frac{(Y - Xb_R)'(Y - Xb_R)}{2\hat{\sigma}_R^4} = 0$$

$$\frac{\partial \ln \mathcal{L}_R(\beta,\sigma^2)}{\partial \lambda} = Rb_R - q = 0$$

- The optimal λ is: $\hat{\lambda} = \sigma^{-2}(R\left(X'X\right)^{-1}R')^{-1}(q - R\hat{\beta})$

$$\begin{aligned}
\text{var}[\hat{\lambda}] &= \sigma^{-4}(R\left(X'X\right)^{-1}R')^{-1}R\text{var}[\hat{\beta}]R'(R\left(X'X\right)^{-1}R')^{-1} \\
&= \sigma^{-2}(R\left(X'X\right)^{-1}R')^{-1}R(X'X)^{-1}R'(R\left(X'X\right)^{-1}R')^{-1} \\
&= \sigma^{-2}(R\left(X'X\right)^{-1}R')^{-1}
\end{aligned}$$

Lagrange Multiplier Test

- The optimal λ is:

$$\hat{\lambda} = \sigma^{-2}(R\left(X'X\right)^{-1}R')^{-1}(q - R\hat{\beta}).$$

- The asymptotic distribution of $\hat{\lambda}$ is

$$\hat{\lambda} \xrightarrow{d} \mathcal{N}(0, \sigma^2(R\left(X'X\right)^{-1}R')^{-1}).$$

- The LM test is

$$\begin{aligned} LM &= \hat{\lambda}'(\hat{\sigma}_R^{-2}R\left(X'X\right)^{-1}R')\hat{\lambda} \\ &= \hat{\sigma}_R^{-2}(R\hat{\beta} - q)'(R\left(X'X\right)^{-1}R')^{-1}(\hat{\sigma}_R^2 R\left(X'X\right)^{-1}R')\hat{\sigma}_R^{-2}(R\left(X'X\right)^{-1}R')^{-1}(R\hat{\beta} - q) \\ &= \frac{(R\hat{\beta} - q)'(R\left(X'X\right)^{-1}R')^{-1}(R\hat{\beta} - q)}{\hat{\sigma}_R^2} \end{aligned}$$

Lagrange Multiplier Test

$$
\begin{aligned}
LM &= \hat{\lambda}'(\hat{\sigma}_R^{-2} R \left(X'X\right)^{-1} R')\hat{\lambda} = \frac{\hat{\lambda}' R \left(X'X\right)^{-1} R' \hat{\lambda}}{\hat{\sigma}_R^2} \\
&= \frac{(R\hat{\beta} - q)'(R\left(X'X\right)^{-1} R')^{-1}(R\hat{\beta} - q)}{\hat{\sigma}_R^2} \\
&= \frac{RSS_R - RSS_U}{RSS_R/n} = n\left(\frac{RSS_R}{RSS_R - RSS_U}\right)^{-1} \\
&= n\left(1 + \frac{RSS_U}{RSS_R - RSS_U}\right)^{-1} \\
&= n\left(1 + \frac{n-k}{JF}\right)^{-1} \simeq JF.
\end{aligned}
$$

Lagrange Multiplier Test

- We could easily see that the three classical tests are asymptotically equivalent.
- In finite sample, the numerical values for the three tests are different.
- One could try to compare the trinity and end up with

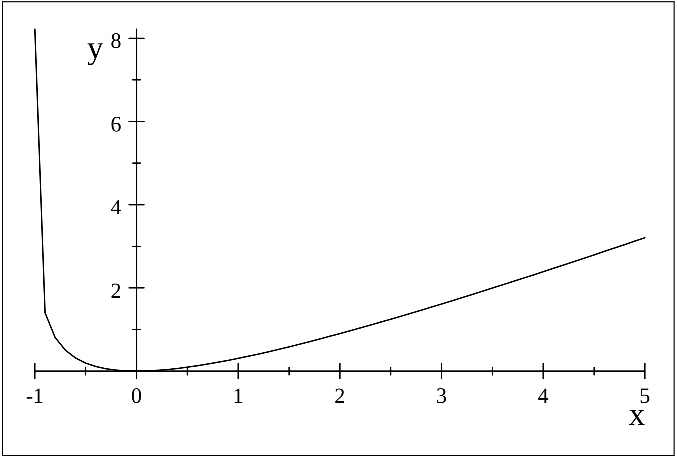
$$W \geq LR \geq LM,$$

where we use the concavity of the log function:

  - $\ln(1+a) \leq a$, and
  - $a < (1+a)\ln(1+a)$.

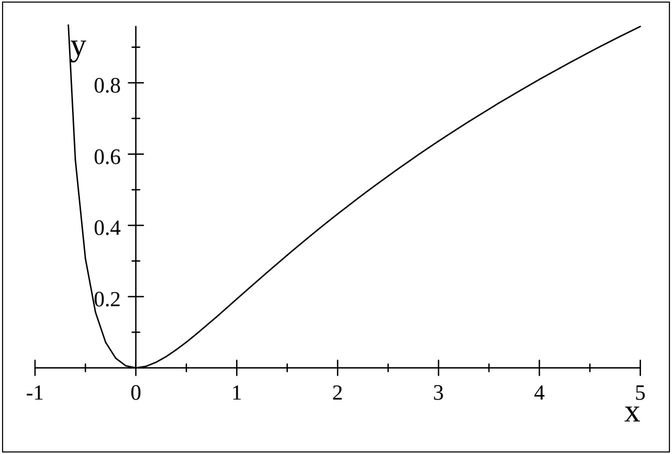
Lagrange Multiplier Test

$$y = x - \ln(1 + x)$$



Lagrange Multiplier Test

$$y = \ln(1+x) - \frac{x}{1+x}$$



Lagrange Multiplier Test

- An alternative view of LM test is to evaluate FOCs of the unrestricted likelihood at the restricted estimates.
- The test is based on the score vector (or gradient). This is the reason why we name the Score test.
- That is,

$$\frac{\partial \ln \mathcal{L}\left(b_R, \hat{\sigma}_R^2\right)}{\partial \beta} = \frac{1}{\hat{\sigma}_R^2} X' \left(Y - Xb_R\right) = \frac{1}{\hat{\sigma}_R^2} X' e_R$$

$$\frac{\partial \ln \mathcal{L}\left(b_R, \hat{\sigma}_R^2\right)}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}_R^2} + \frac{1}{2\hat{\sigma}_R^4} \left(Y - Xb_R\right)' \left(Y - Xb_R\right) = 0.$$

Lagrange Multiplier Test

- The score evaluated at restricted estimates is

$$s\left(\theta_R\right) = \left(\begin{array}{c} \frac{1}{\hat{\sigma}_R^2} X' e_R \\ 0 \end{array} \right).$$

- The variance of the score is simply the information matrix.

$$\mathcal{I}(b_R, \hat{\sigma}_R^2) = \left[\begin{array}{cc} X'X \times \hat{\sigma}_R^{-2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}_R^4} \end{array} \right].$$

Lagrange Multiplier Test

- Under the null, one could form a quadratic form to conduct the testing.

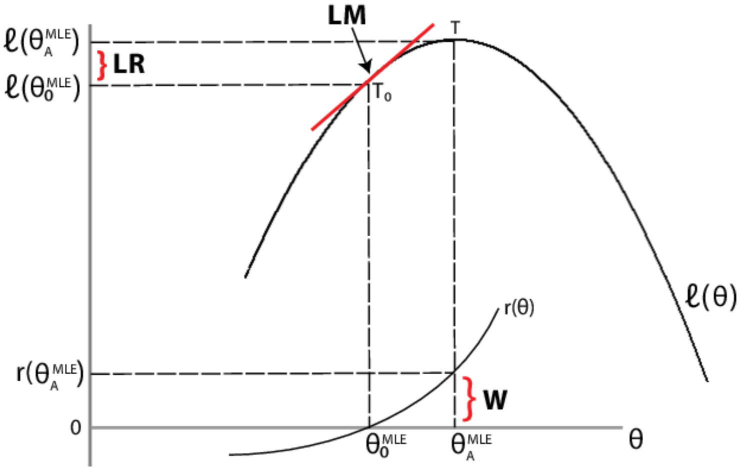
$$
\begin{aligned}
LM &= s(\theta_R)' \mathcal{I}(b_R, \hat{\sigma}_R^2)^{-1} s(\theta_R) \\
&= \frac{e_R' X (X'X)^{-1} X' e_R}{\hat{\sigma}_R^2} \\
&= n \frac{e_R' X (X'X)^{-1} X' e_R}{e_R' e_R} \\
&= nR^2.
\end{aligned}
$$

- The R^2 is the uncentered R-square calculated from the regression, $e_R = X\gamma + \eta$. However, if the restriction does not involve intercept term (say, $\beta_2, \beta_3, ..., \beta_k$), uncentered R-square will coincide with the centered R-square.

Lagrange Multiplier Test

- Note that $X'e_R = R'\hat{\lambda} \sim \mathcal{N}(0, \sigma^2 R'(R\,(X'X)^{-1}\,R')^{-1}R)$
- or $X'e_R = R'\hat{\lambda} \overset{d}{\sim} \mathcal{N}(0, \sigma^2 R'(R(\text{plim}X'X/n)^{-1}R')^{-1}R)$
- This has a singular variance covariance matrix $R'(R\,(X'X)^{-1}\,R')^{-1}R$
- Generalized inverse would be needed! (See Greene!)

Graphical Illustration of Three Tests



Test for Non-linear Restriction

- It will be relatively easy to test non-linear restriction using Wald test since it may be difficult to impose nonlinear restrictions in estimation.
- Assume the nonlinear restriction is $g(\beta) = q$. Using the multivariate delta method, the Wald test statistic is

$$W = (g(\hat{\beta}) - q)' \left(\frac{\partial g(\hat{\beta})}{\partial \beta}' \text{var}(\hat{\beta}) \frac{\partial g(\hat{\beta})}{\partial \beta} \right)^{-1} (g(\hat{\beta}) - q) \overset{d}{\sim} \mathcal{X}^2 (J),$$

where J is typically the number of nonlinear restrictions.

Non-invariance of the Wald test

- Drawback of Wald type test for nonlinear restriction?
- Testing $H_o : \theta = 0$ vs. $H_o : \theta^3 = 0$
- Formulating the nonlinear hypothesis, say $\beta_1\beta_2 = 1$ or $\beta_1 = 1/\beta_2$
- How about LR and LM?
    - invariant to reformulating the restrictions

Non-invariance of the Wald test: Example

Example

Consider the model

$$Y_i = \beta + \varepsilon_i \quad \text{and} \quad H_o : \beta = 1$$

The standard t test is

$$W = \frac{\hat{\beta} - 1}{\sqrt{\hat{\sigma}^2/n}},$$

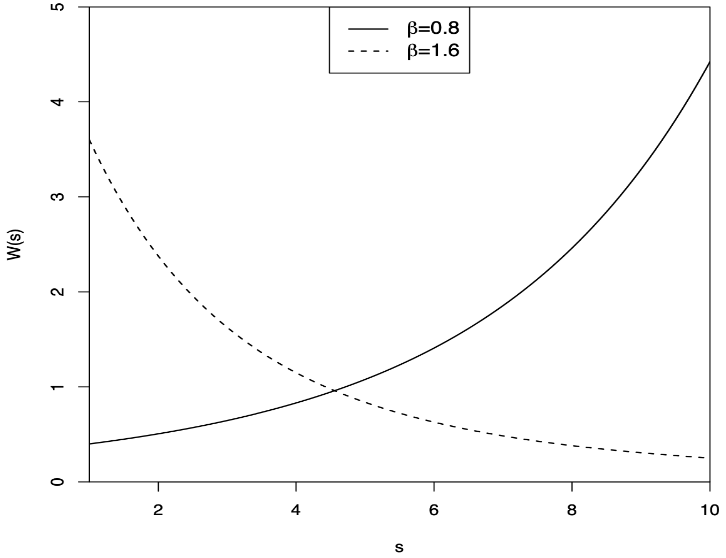
The standard Wald test is

$$W = (\hat{\beta} - 1)(\frac{\hat{\sigma}^2}{n})^{-1}(\hat{\beta} - 1) = n\frac{(\hat{\beta} - 1)^2}{\hat{\sigma}^2},$$

Now consider an equivalent test as follows:

$$H_o(s) : \beta^s = 1$$

Non-invariance of the Wald test: Graph



References

1. Davidson, R. and J. G. MacKinnon (1993) *Estimation and Inference in Econometrics*. 2nd ed. New York: Oxford University Press.
2. Greene, W. H. (2008) *Econometric Analysis*, 6th edition, Prentice Hall.
3. Johnston and DiNardo (1997) *Econometric Methods*, 4th edition, McGraw Hill.